AD-A118 730 PORTLAND STATE UNIV OR DEPT OF PSYCHOLOGY F/G 5/10
MODELS AND ESTIMATION PROCEDURES FOR THE ANALYSIS OF SUBJECTS-B--ETC(U)
JUN 82 J A PAULSON N00014-79-C-0214
UNCLASSIFIED TR-82-1 NL

1 OF 1
AD A
118730

END
DATE
FILMED
09-82
DTIC

AD A118730

(9)

Technical Report 82-1
June 1982

MODELS AND ESTIMATION PROCEDURES FOR
THE ANALYSIS OF SUBJECTS-BY-ITEMS
DATA ARRAYS: THE FINAL REPORT

James A. Paulson
Psychology Department
Portland State University

This research was supported by the Defense Advanced Research Projects
Agency and monitored by the Office of Naval Research, Personnel and
Training Research Programs, under Contract N00014-79-C-0214, NR 154-429.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report 82-1	2. GOVT ACCESSION NO. AD A [illegible]	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Models and Estimation Procedures for the Analysis of Subjects-by-items Data Arrays: The Final Report		5. TYPE OF REPORT & PERIOD COVERED FINAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James A. Paulson		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C0214
9. PERFORMING ORGANIZATION NAME AND ADDRESS Psychology Department Portland State University Portland, OR 97207		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N RR 042-06-01 NR 154-429
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22217		12. REPORT DATE June 30, 1982 13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Transformations
Nonadditivity
Empirical Bayes estimation
Unidimensionality
Finite state latent spaces
Signed-number arithmetic test performance

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This final report describes results obtained in a study of models and estimation procedures for the analysis of subjects-by-items data arrays. The results include a new transformation to render an array additive and some simple empirical Bayes methods for estimating subject and item marginal effects. Unfortunately, the problems which these results concern are not as central to the problems of model-based psychological measurement as I originally thought. The report discusses a new problem which I began working on roughly midway through this project, a problem which is crucial for the foundations of model-based mastery testing.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MODELS AND ESTIMATION PROCEDURES FOR THE ANALYSIS OF SUBJECTS-BY-ITEMS DATA ARRAYS: THE FINAL REPORT

Overview

The problems suggested by the title of the project which is the subject of this report are of intrinsic interest from the point of view of mathematical statistics and scaling theory. Unfortunately, they turn out not to be as central to the problems of model-based psychological measurement as I originally thought. Thus, roughly midway through the project I started to focus on a new problem which I feel is more central to the concerns of this program and deserves much more attention than it has received so far. In this report, I will summarize the progress made on the original problems, describe some of the unresolved difficulties, and explain why it is important for the advancement of model-based psychological measurement to address the new problem mentioned above.

The problems originally considered concern estimation of parameters in Tukey's (1949) model for data in a two-way array. This generalization of the additive model is interesting for a couple of reasons. First, data arrays which are nonadditive, but can be rendered additive by a monotonic transformation, often fit this model very well. Second, the model has some of the same advantages as the additive model for purposes of scaling row and column effects (subject ability and item difficulty effects, in the present application). Row means are independent of the column parameters and vice versa, so row and column means can provide a suitable basis for scaling the row and column effects. For these reasons, Tukey's model is worth considering as an alternative to Rasch models as a basis for scaling unidimensional latent traits.

If one adopts Tukey's model as the basis for testing, there are two ways

one might go about measuring subject and item effects, assuming the original data are nonadditive. One might seek a transformation that will render the data additive. To the extent that such a transformation brings the array more into line with the usual ANOVA model assumptions, there may be advantages to carrying out estimation in terms of the transformed data and then transforming the results back into the original metric. On the other hand, it is sometimes undesirable or useless to transform the original data. For example, transformation of dichotomous test data is useless. As has been mentioned, row and column means of the original data provide a reasonable basis for measurement of row and column effects, without using a transformation. The work of this project has resulted in two contributions that should help in either of these two approaches.

If Tukey's model holds and if it is possible to render the data additive by a monotonic transformation, then the required transformation is a simple one implicit in the model. It is unique up to a linear transformation. Let y represent data and let μ be the grand mean of the array and λ be the nonadditivity parameter in the model, and let $c=\mu-\lambda^{-1}$. The data cannot be rendered additive by <u>any</u> monotonic transformation unless either $E(y)>c$ for all data points, or $E(y)<c$ for all data points. If either condition is satisfied, then the required transformation to attain additivity is $\log(y-c)$ or $\log(c-y)$, depending on which condition holds. It is remarkable that this result has not been noted before, considering the fact that Tukey's model is so well known. Tukey's model and results concerning it are summarized in more detail in the next section. A complete account is given in Technical Report 80-1.

The problem of estimation of subject and item effects arises whether or not one transforms the data first. The usual least squares estimators are unbiased and consistent, but they can be improved upon using an empirical Bayes approach.

Griffin and Krutchkoff (1971) have shown that the optimal linear estimators of marginal effects are the sample effects multiplied by a constant. The constant is equal to the ratio of the variance of the underlying parameter to the sum of that variance and the conditional variance of the sample effect, given the parameter. A very important feature of their result for our purposes is that it does not depend on parametric assumptions concerning the distribution of entries in the array. It applies equally well to dichotomous, right-wrong responses and to response latencies, for example. I have shown how this approach can be applied to the problem of estimation of several proportions. Jackson (1972) suggested a method similar to the one I propose, but Novick, Lewis, and Jackson (1973) found some difficulties with his suggestion. These difficulties lead them to propose a more complicated approach. A key problem in applying either the Griffin and Krutchkoff approach or Jackson's approach to this problem is estimation of the variance of the underlying parameter. I propose a weighted estimator of this variance and have shown that, in the context of estimation of several proportions, it avoids the serious problems noted by Novick, Lewis, and Jackson. These developments are described in detail in Technical Report 81-1, which will be summarized below, in the third section of this report.

The research just described assumes one is dealing with unidimensional traits. There is no good reason to assume a priori that a unidimensional latent trait is capable of representing the possible states of learning of subjects with regard to a given conceptual domain. Unidimensional latent trait models are postulated on the basis of mathematical convenience. Unfortunately, there are numerous situations in achievement testing where data show that unidimensionality does not hold.

Failure of unidimensionality can have serious practical consequences in adaptive mastery testing, one of the areas where it has been hoped that latent

trait theory might prove most beneficial. If a given unidimensional model holds, it is possible to determine the status of subjects to a given level of precision with substantially fewer items than a standard test would require. This is accomplished by taking the subject's responses to the initial items on a test into account in selecting subsequent items. However, if the model on which item selection is based is wrong, then the adaptive procedure may introduce biases and other inaccuracies into the assessment of the subject which make it worse than a test constructed on traditional principles, rather than better.

Since the success of adaptive testing may well hinge on the quality of the models it is based on, a thorough exploration of alternative models to represent the state of subjects in different cognitive domains should be a first priority for future research. Such work should draw on recent developments in cognitive psychology, but it will need to go considerably beyond them. As Susan Whitely (1980) has pointed out, these recent developments have been preponderantly concerned with chronometric studies of the performance of subjects who are competent at the tasks they are asked to do. Dealing with the testing problem will require models that represent the state of subjects who are still learning the tasks in question and can account for the subjects' patterns of correct and incorrect responses. Multicomponent latent trait models represent one approach to the problem which is worth pursuing. An even more promising approach, in my opinion, is to use a finite, discrete state latent space to represent the state of the subject. The all-or-none models of concept learning popular in the 1960's are the sort of thing I have in mind, but one must go considerably beyond those models to represent the state of subjects with respect to concepts one would be interested in testing in practical situations. Much of my time in the latter part of this project was spent developing a proposal outlining how these ideas might be used in testing mastery of signed-number arithmetic. Since this approach is

now being explored on a new contract, detailed discussion of it will be deferred to subsequent reports.

<u>Tukey's model and its properties</u>

In Tukey's two-way ANOVA model with a single-degree of freedom for non-additivity the expected value of the score of subject i on item j is

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j \ . \tag{1}$$

The linear transformation obtained by multiplying both sides of Equation 1 by λ and adding $1 - \lambda\mu$ yields a multiplicative form of the model:

$$\begin{aligned} \lambda\mu_{ij} - \lambda\mu + 1 &= 1 + \lambda\alpha_i + \lambda\beta_j + \lambda^2\alpha_i\beta_j \\ &= (1 + \lambda\alpha_i)(1 + \lambda\beta_j) \ . \end{aligned} \tag{2}$$

If $1 + \lambda\alpha_i > 0$ for all i and $1 + \lambda\beta_j > 0$ for all j, then we have the additive representation

$$\log(\lambda\mu_{ij} - \lambda\mu + 1) = \log(1 + \lambda\alpha_i) + \log(1 + \lambda\beta_j) \ . \tag{3}$$

A theorem of Luce and Tukey (1964) implies that this additive representation is unique up to a linear transformation, when it exists. Equation 3 implies that the transformation $\log(\lambda y - \lambda\mu + 1)$ should yield additivity for an array of data satisfying Tukey's model, provided one can take logarithms on both sides of Equation 2. Obvious linear transformations yield more convenient, equivalent forms of the transformation: $\log(y - \mu + \lambda^{-1})$ if λ is positive, $\log(\mu - \lambda^{-1} - y)$ if λ is negative. It is interesting and somewhat puzzling that Tukey never mentions this transformation explicitly in connection with his model. He does mention transformations of the form $\log(y - c)$ in passing, but not the important point that $c = \mu - \lambda^{-1}$. The role of c in these transformations is analogous to

the role of the exponent in power transformations, so the fact that c is determined by the model parameters is significant.

The number $\mu - \lambda^{-1}$ has a geometric interpretation which shows an interesting aspect of Tukey's model. Notice that for a given subject i, the regression of the expected score on the β_j's is linear, with slope $1 + \lambda\alpha_i$. If we omit the subscript on β in Equation 1 to indicate that it is variable and write $\mu_i(\beta)$ in place of μ_{ij}, we obtain

$$\mu_i(\beta) = \mu + \alpha_i + (1 + \lambda\alpha_i)\beta \ . \tag{4}$$

Solving for the point of intersection of any two distinct regression lines given by Equation 4 yields $\beta = -\lambda^{-1}$, at which point the expected score is $\mu_i(-\lambda^{-1}) = \mu - \lambda^{-1}$. Thus, in graphical terms, Tukey's model posits linear regression lines for each subject which converge to a common focal value. The expected score at that value is $\mu - \lambda^{-1}$.

It is easy to verify the following properties of deviations from the focal value:

1. The overall mean deviation is λ^{-1}.
2. The ratio of the mean deviation for subject i to the overall mean deviation is $1 + \lambda\alpha_i$.
3. The ratio of the mean deviation for item j to the overall mean deviation is $1 + \lambda\beta_j$.
4. The expected deviation from the focal value for subject i on item j is equal to the product of the mean deviation for subject i and the mean deviation for item j, divided by the overall mean deviation.

If $\mu = \lambda^{-1}$, the focal value is zero and the remarks above hold for the scores themselves. The last remark can then be strengthened to say that the expected score for subject i on item j is equal to the product of the mean score for

subject i and the mean score on item j, divided by the overall mean score. This is the multiplicative property which is the foundation of Rasch's (1959) Poisson and gamma models. Rasch's model for analysis of dichotomous items has the same multiplicative structure if we regard the dependent variable to be the <u>odds</u> in favor of correct response. Thus, Tukey's model is seen to be a direct generalization of the Rasch models.

The generalization of Rasch models represented by Tukey's model costs less in loss of desirable properties than one might expect. One obtains in return a model with a much larger range of applicability. It is true that the property of separability of item and subject parameters which Rasch has emphasized does not hold in the strict sense employed by Rasch. Item parameters are separable from subject parameters in Rasch's sense if the conditional probability of a subject making a given response to an item, given the subject's total score on all items, depends only on the item parameters, not on the subject parameter. This is not the case in the general model, but the item parameters are nevertheless separable from the subject parameters in a weaker sense that is useful. The deviation of the mean score on an item from the overall mean depends only on the item parameter, not on the subject parameters. In fact, the usual least squares estimators of subject and item effects for the two-way ANOVA model are unbiased and consistent.

It is easy to verify that item and subject parameters are not separable in Rasch's sense, even if we restrict ourselves to an additive model with standard normal distribution assumptions. This is the best behaved model that one can assume in terms of how well it lends itself to estimation of parameters. The fact that its item and subject parameters are not separable in Rasch's sense suggests that Rasch's sense of separability is an unnecessarily restrictive requirement to place on models for subjects-by-items arrays.

Some interesting properties of the Rasch models do carry over to the more general model if one assumes, as Rasch did, that the conditional distribution of responses, given the subject and item parameters, is a Poisson or a gamma distribution. Due to the special reproductive properties of these distributions, the marginal sums over subjects or items are distributed as Poisson or gamma variates; i.e., the form of the conditional distribution carries over to the marginal distribution. Furthermore, the parameters of the distribution of the marginal sums for each student depend only on the student parameters and the parameters of the distributions of the marginal sums for each item depend only on the item parameters. Thus, subject and item parameters are separable in Tukey's model in senses which are important for the problem of parameter estimation. We turn now to this problem.

<u>Estimation of the parameters in Tukey's model</u>

It has already been noted that deviations of subject and item marginal means from the grand mean are consistent, unbiased estimators of subject and item effects in Tukey's model. Recent work on empirical Bayes estimation procedures suggests ways to improve upon the least squares estimators. Previous work has shown that the usual estimator of the nonadditivity parameter in Tukey's model can be severely biased. Because the usual estimator depends on estimates of the marginal effects, it seemed that the corrections to the least squares estimators of marginal effects which the empirical Bayes estimators make might lead to a corresponding correction in the bias of the nonadditivity parameter estimator. This project has explored these questions in detail.
The study of the empirical Bayes estimators of subject and item effects has led to some useful results, but carryover of improvement to the estimator of λ has not worked out as hoped.

<u>Estimation of subject and item effects</u>. Let X be a random variable with expected value $\underline{m}$. Suppose we have a random sample of $\underline{n}$ different X's from a population in which the mean $\underline{m}$ is not constant, but has a distribution with mean μ_m and variance σ_m^2. Let σ_x^2 be the unconditional variance of X. In the present context, X might be a marginal mean for a given subject, $\underline{m}$ the true mean for the subject, σ_m^2 the true variance in $\underline{m}$ among subjects, and σ_x^2 the variance of the sample marginal means for subjects.

The question arises about how best to estimate the separate means using the values of the separate X's obtained. One approach would be to use each X to estimate the mean in its subpopulation. If the X's are subject means, for example, we would use the subject means themselves to estimate the subjec true scores. On the other hand, if the sample variance of the X's observed i close to what one would expect in sampling from a population with a single, constant value of $\underline{m}$, it might be better to use the grand mean of the X's as a common estimate of $\underline{m}$ for all the means. Griffin and Krutchkoff (1971) have shown that if one restricts oneself to estimators which are linear functions the X's, the best estimators in the sense of minimizing overall squared error are given by the following compromise between the possibilities just mentioned:

$$\begin{aligned} \hat{\underline{m}}(X) &= CX + (1 - C)\mu_m \\ &= \mu_m + C(X - \mu_m), \end{aligned} \tag{5}$$

where $C = \sigma_m^2 / \sigma_x^2$. The two possibilities mentioned above correspond to letting $C = 0$ if $\sigma_m^2 = 0$, and $C = 1$ if $\sigma_m^2 = \sigma_x^2$.

Implementation of the optimal linear estimators requires knowledge of μ_m and σ_m^2 which we usually lack. The sample grand mean is a good estimator of μ_m, but the estimation of σ_m^2 can be something of a problem. The standard estimator

of σ_m^2 from the analysis of variance components is probably satisfactory for present purposes, but it cannot be claimed that the estimators of marginal effects which result from using these estimates of μ_m and σ_m^2 are optimal. They do seem to be an improvement on the least squares estimators.

Estimation of the nonadditivity parameter. Let y_{ij} be the score of subject i on item j and let $\hat{\alpha}_i$ and $\hat{\beta}_j$ be the least squares estimators of the subject and item effects for subject i and item j, respectively. The usual estimator of the nonadditivity parameter λ is given by

$$\hat{\lambda} = \frac{\sum_{i,j} \hat{\alpha}_i \hat{\beta}_j y_{ij}}{\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2} . \tag{6}$$

The expected value of the numerator of the right side of Equation 6 is $\lambda \sum_{i,j} \alpha_i^2 \beta_j^2$. The denominator is an estimator of $\sum_{i,j} \alpha_i^2 \beta_j^2$, so $\hat{\lambda}$ is a plausible estimator of λ. Unfortunately, while $\hat{\alpha}_i$ and $\hat{\beta}_j$ are unbiased estimators of α_i and β_j, $\sum_i \hat{\alpha}_i^2$ and $\sum_j \hat{\beta}_j^2$ tend to overestimate the respective sums of squared effects. Their expectations involve σ_e^2 as well as σ_α^2 and σ_β^2. As a result, $\hat{\lambda}$ tends to underestimate λ in absolute value. The bias can be substantial in realistic settings. The median of $\hat{\lambda}$ is approximately

$$\text{median } \hat{\lambda} \doteq C_\alpha C_\beta \lambda, \tag{7}$$

where

$$C_\alpha = \frac{J\sigma_\alpha^2}{\sigma_e^2 + J\sigma_\alpha^2} \quad \text{and} \quad C_\beta = \frac{n\sigma_\beta^2}{\sigma_e^2 + n\sigma_\beta^2} ; \tag{8}$$

Note that C_α and C_β are the correction factors by which the optimal linear estimators of Griffin and Krutchkoff would correct the least squares estimators of

subject and item effects, respectively

Equation 7 suggests that we might improve on $\hat{\lambda}$ by multiplying by $C_\alpha^{-1}C_\beta^{-1}$. This would undoubtedly help if we knew C_α and C_β, but we do not know them in most situations, so we must estimate them. We did simulation studies which show that multiplying $\hat{\lambda}$ by analysis-of-variance components estimators of C_α^{-1} and C_β^{-1} does reduce bias, but at a disastrous cost in variability. The resulting estimator has substantially greater squared error than $\hat{\lambda}$. A good estimator of λ is yet to be found.

It is important to put the difficulties in the estimation of λ in proper perspective. The usual estimator is consistent, which is about as much as can be claimed for the available estimators of parameters in competing logistic models. The fact that the estimation of λ poses some problems should not in itself lead one to dismiss the Tukey model as an attractive alternative to Rasch models. For one thing, the difficulties do not affect the estimation of the parameters of most concern, the subject and item parameters. On the other hand, the uncertainty regarding λ would affect the application of Tukey's model in adaptive testing.

Suppose one is trying to estimate α_i for some subject on the basis of the subject's performance on a small subset of the items. Let P_i be the subject's average score on the subset of items and let $\bar{\beta}$ be the average item effect for items in the subset. Then the method of moments estimator of α_i is

$$\tilde{\alpha}_i = \frac{P_i - \mu - \bar{\beta}}{1 + \lambda\bar{\beta}} \quad (9)$$

Clearly, the value of λ would affect judgements about α_i. If all items were employed, $\bar{\beta}$ would be zero and $\tilde{\alpha}_i$ would not be affected by λ, but one of the

main points to adaptive testing is to tailor the items to either the ability of the subject or the ability level adopted as a cutoff, depending on the testing context. In either case, $\bar{\beta}$ is unlikely to be zero. The method of moments estimator is probably not the best we can do in this situation, but whatever approach is employed would be affected by λ, so it would be desirable to have a better estimator of this parameter.

Empirical Bayes estimation of proportions in several groups. The optimal linear estimators of Griffin and Krutchkoff can be applied to the problem of simultaneous estimation of proportions in several groups. In order to do so it is necessary to estimate the variance of the underlying proportions. Griffin and Krutchkoff do not deal with the unbalanced case, where the observed proportions are based on different sample sizes. In Technical Report 81-1, I propose the following estimator for σ_m^2, the between-group variance component in the unbalanced one-way random effects ANOVA model:

$$\hat{\sigma}_m^2 = (J - 1)\left(N - \frac{\Sigma n_j^2}{N}\right)^{-1} \left(MS_{between} - MS_{within}\right) . \qquad (10)$$

Here J is number of groups, n_j the number of observations in group j, and N is the total number of observations. This estimator is an unbiased, consistent estimator of σ_m^2 even if the usual assumptions of normality and homoscedasticity fail to hold. When Equation 10 is specialized to the case of binomial proportions, the formulas for the mean sums of squares can be simplified to expressions involving the numbers of successes in each group, r_j's, and the total number of successes, R. The estimator of the variance of the underlying proportions becomes

$$\hat{\sigma}_p^2 = \frac{\frac{N(N-1)}{N-J} \sum_j \frac{r_j^2}{n_j} - R^2 - \frac{N-1}{N-J} R}{N^2 - \sum_j n_j^2} . \qquad (11)$$

The optimal linear estimator of each of several proportions corresponds to classical Bayes estimation, using the means of the posterior distributions of p as point estimates. One starts with a beta prior distribution on p having parameters

$$n' = \frac{\mu_p(1-\mu_p) - \sigma_p^2}{\sigma_p^2} \text{ and } r' = n'\mu_p \ , \tag{12}$$

where μ_p and σ_p^2 are the mean and variance of the prior distribution. The posterior means are given by

$$\hat{p}_j = \frac{r_j + r'}{n_j + n'} \ . \tag{13}$$

In practice, we estimate n' and r' by substituting in Equation 12 the estimate

$$\hat{\mu}_p = \frac{R}{N} \tag{14}$$

for μ_p and $\hat{\sigma}_p^2$ from Equation 11 for σ_p^2. Then we apply Equation 13 to the data for each successive group.

The approach to estimation of several proportions just described is extremely easy to implement. It is interesting to compare it with another very straightforward approach due to Jackson (1972). Jackson's method is similar in spirit to ours. The main differences are that it deals with root arcsine transformations of the proportions rather than dealing directly with the proportions and that it employs unweighted estimators of the mean and variance of the underlying distribution of transformed proportions, rather than weighted estimators.

Novick, Lewis, and Jackson (1973) applied Jackson's approach to four sets

of data gathered in educational contexts. In several of the applications, they found the approach less than satisfactory. The main problem seems to be the sampling variability of the estimator of the variance of the underlying parameter. For example, a negative estimate of variance occurs in one of the applications. Novick et al. argue that the difficulties with estimation of the between-group variance component show the need for a more rigorously Bayesian approach in which prior beliefs about the extent of between-group variation can be integrated with sample information.

When Griffin and Krutchkoff's method is applied to the data of Novick et al., with weighted estimates of σ_p^2, the problems encountered by Jackson's approach do not arise. The resulting estimates are in line with some of the Bayesian solutions considered by Novick et al. This suggests that the use of unweighted estimators might be the root of the problem with Jackson's approach, rather than the failure to incorporate prior beliefs into the estimates.

Let g_j denote the root-arcsine transformation of the observed proportion of successes in group j:

$$g_j = \sin^{-1}\left(\frac{r_j + 3/8}{n_j + 3/4}\right)^{\frac{1}{2}} . \tag{15}$$

Let γ be the transformed value of a proportion p. Jackson estimates the mean and variance of the distribution of γ by

$$\tilde{\mu}_\gamma = \frac{\sum_j g_j}{J} \tag{16}$$

and

$$\tilde{\sigma}_\gamma^2 = (J-1)^{-1} \sum_j (g_j - \tilde{\mu}_\gamma)^2 - J^{-1} \Sigma(4n_j+2)^{-1}, \tag{17}$$

respectively. The conditional variance of g_j is $(4n_j + 2)^{-1}$, so $\tilde{\sigma}_Y^2$ is the difference between the sample variance of the g_j's and the average of their conditional variances. The following estimators of μ_Y and σ_Y^2 weigh the contributions from each group inversely to the conditional variance of g_j:

$$\bar{g}. = (4N + 2J)^{-1} \sum_j (4n_j + 2)g_j \ , \tag{18}$$

$$\hat{\sigma}_Y^2 = \left(4N + 2J - \frac{\Sigma(4n_j + 2)^2}{4N + 2J}\right)^{-1} \left(\Sigma(4n_j + 2)(g_j - \bar{g}.)^2 - (J - 1)\right) .$$

When Jackson's method is applied to the data of Novick et al. using the weighted estimators in Equation 18 instead of the unweighted estimators in Equations 16 and 17, the results are practically identical to the Griffin and Krutchkoff estimates. On the face of it, weighted estimators would seem to be preferable to unweighted estimators, but under some circumstances one would be too hasty in drawing this conclusion. Both weighted and unweighted estimators can yield negative estimates of the between-group variance component and Tukey (1957) has shown that there are situations in the unbalanced case where unweighted estimators are better than weighted.

According to Tukey (1957), two of the most important factors determining whether weighted or unweighted estimators will yield more accurate estimates of the between-group variance component are the size of true between-group variance and the variability of sample size from group to group. Other things being equal, large true variation between groups favors unweighted estimators over weighted estimators, whereas large variability in sample size favors weighted estimators. In Monto Carlo simulations described in Technical Report 81-1, under conditions mimicking those in the examples analyzed by Novick et al.,

we found that the differences in performance between weighted and unweighted estimators of the between-group variance are not significant when the true variation between groups is large and the variability in sample size is moderate; that is, under conditions that favor unweighted estimators. On the other hand, the differences in favor of weighted estimators are substantial when the sample sizes are extremely variable and the between-group variance component is small. It is therefore recommended that weighted estimators be used as a general practice when applying either the Griffin and Krutchkoff or the Jackson procedure.

From a philosophical point of view, some might prefer the rigorously Bayesian approach of Novick et al. to the empirical Bayes approaches of Griffin and Krutchkoff or Jackson. However, in applying the rigorously Bayesian approach there are problems of computational complexity and technical problems concerning specification of prior distributions which might lead even a convinced Bayesian to prefer the empirical Bayes approaches. These problems are discussed in detail in Technical Report 81-1. The work on this project shows that the use of weighted estimators of the between-group variance component answers the most serious objections of Novick et al. to the empirical Bayes approach.

<u>Conclusions and recommendations</u>

The usefulness of Tukey's model for model-based psychological testing is probably greatest for analyses of responses which are not dichotomous, such as response latencies. There are not many models for such responses to choose from, other than the additive ANOVA model and Rasch's multiplicative models. Tukey's model is a generalization of these models which costs very little in terms of loss of their desirable properties. Nondichotomous variables have not traditionally been the focus of much attention in test development. However,

recent efforts, such as Sternberg (1977) and Thissen (1979), to relate the chronometric methods of analysis frequently employed in the study of cognition to psychometrics, suggest that the analysis of response latencies will be of increasing importance in the future. The results of this project should be helpful in these analyses.

When the response variable is dichotomous, Tukey's model may or may not be competitive with Rasch models and other logistic models. Whether it is or not undoubtably depends on the specific test in question. Tukey's model is most likely to be adequate in testing situations with the following characteristics:

1. The items are unidimensional.
2. The Rasch model is not appropriate for some reason, such as unequally discriminating items.
3. The number of subjects is too small to adequately estimate the parameters in the more flexible two- or three-parameter logistic models.

Under these circumstances, Lord (1979) has shown that the use of Rasch methods is preferable to methods based on the two- or three-parameter logistic models, even when a two- or three-parameter logistic model generates the data. Although the studies comparing the performance of methods based on Tukey's model with the performance of the Rasch methods have not been done, it seems plausible that Tukey's model would often provide a better approximation than the Rasch methods when the conditions listed above hold.

Unfortunately, these conditions define an uncomfortably narrow range of applicability for Tukey's model--narrower than is perhaps apparent on first reading. It is not hard to find situations satisfying the second and third conditions. The Rasch model is often too simple to fit the data, and Lord's results suggest that it is impractical to use more complex logistic models unless there are at least two or three hundred subjects available for item

calibration. The unidimensionality condition is the hard one to satisfy. How hard it is likely to be is not sufficiently appreciated yet by workers in the field.

A study of Tatsouka and Birenbaum (1979) of the performance of prealgebra students on a test of signed-number arithmetic illustrates a problem which makes the achievement of unidimensionality problematic. The students all took computer-assisted drills on signed numbers but their instruction in class varied from teacher to teacher. If their test performance could be adequately represented by a unidimensional latent trait model, then item characteristic curves based on different samples of subjects ought to coincide. Tatsuoka and Birenbaum found dramatic differences between the item characteristic curves for different classes, differences which they could relate plausibly to specific item types and differences in the approaches taken by the various classroom teachers in presenting the signed-number concept.

Failures of unidimensionality can often be attributed to the inclusion of items in a test which clearly tap different skills. This argument does not apply in the present case. The test of signed-number arithmetic was very carefully designed to represent just about as precisely and narrowly defined a domain as one could have and still have a testing situation of genuine applied interest. Upon reflection, it should not be too surprising to find that unidimensionality is difficult to achieve with such tests. Most achievement tests are intended to assess knowledge of material from a specified domain, without reference to how the material was taught. It is reasonable to suppose that the relative difficulty of test items depends on how the concepts which the items are designed to test were taught and learned. Due to variation in instruction, which will usually occur to some extent, even in contexts where student progress is monitored automatically, examinees will differ with respect to how much the

material to be tested is studied and how the study time is used. As a result we should often expect to find the variation from class to class in item characteristic curves which Tatsuoka and Birenbaum found with signed-number arithmetic. Unidimensionality is likely to be an exceptional occurence in achievement tests of narrowly defined domains.

Violation of the unidimensionality assumption can have serious consequences for adaptive tests of concept mastery. Any adaptive procedure based on a unidimensional latent trait model would result in tests comprised of decidedly nonrandom samples of items from the item pool. If the model is incorrect, then the procedure is quite likely to be inappropriate and the resulting measurements misleading.

When I undertook this project, I tacitly assumed that many mastery testing situations would satisfy the conditions under which Tukey's model would be likely to be of most use. I was not thinking enough about the unidimensionality assumption which Tukey's model shares with most of the other extant latent trait models. The areas where unidimensional representations are inadequate are areas of considerable importance in military training and education generally. Many critical specialized training modules involve concepts analogous to signed-number arithmetic. There must be better ways to conceive of mastery of these concepts than to think of it in terms of a subject's position on a continuous latent trait. Finding adequate representations of the performance of subjects on tests of these concepts should have at least equal priority with further development of unidimensional models.

A good clue about a possible direction to pursue in seeking simple models of concept mastery is implicit in the observations of Tatsuoka and Birenbaum and in closely related work of Brown and Burton (1978). Careful analyses by these investigators of the patterns of responses to items on tests of simple

mathematical concepts show that subjects who have not mastered a concept often fall into distinctive patterns of errors, patterns which derive from systematic misconceptions. For example, on signed-number addition problems, some students get the sign right on all the problems, but always add the absolute values of the addends to get the absolute values of the sums, rather than subtracting the absolute values when the signs of the addends disagree. If a significant fraction of the subjects have this misconception, that in itself leads to a violation of unidimensionality. The fact that relatively few systematic misconceptions can account for a majority of the response patterns of subjects who have not mastered a concept suggests an alternative model to serve as a basis for testing mastery of the concept. Instead of characterizing subjects by positions on a numerical continuum, one can characterize them as belonging to latent states. The latent states correspond to the systematic misconceptions, or to mastery of the concept. The development of models along these lines is now the goal of my research in this area.

References and reference notes

Brown, J.S. and Burton, R.R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Griffin, B.S. and Krutchkoff, R.G. Optimal linear estimators: An empirical Bayes version with application to the binomial distribution. Biometrika, 1971, 58, 195-201.

Jackson, P.H. Simple approximations in the estimation of many parameters. British Journal of Mathematical and Statistical Psychology, 1972, 25, 213-228.

Lord, F.M. Small N justifies Rasch methods. Paper presentation at the 1979 Computerized Adaptive Testing Conference in Minneapolis, Minnesota, June 1979.

Luce, R.D. and Tukey, J.W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

Novick, M.R., Lewis, C., and Jackson, P.H. The estimation of proportions in m groups. Psychometrika, 1973, 38, 19-46.

Paulson, J.A. A transformation yielding an additive representation of data in a two-way array. Technical Report 80-1. Portland State University, August 1980.

Paulson, J.A. Empirical Bayes estimation of proportions in several groups. Technical Report 81-1. Portland State University, October 1981.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson and Lydiche (for Danmarks Paedagogiske Institut), 1959.

Sternberg, R. Intelligence, information processing, and analogical reasoning: the componential analysis of human abilities. Hillsdale, N.J.: Erlbaum, 1977.

Tatsuoka, K. and Birenbaum, M. The danger of relying solely on diagnostic adaptive testing when prior and subsequent instructional methods are different. Computer-based Education Research Laboratory Report E-5. University of Illinois, March 1979.

Thissen, D. Latent trait scoring of timed ability tests. In Weiss, D.J. (ed.) Proceedings of the 1979 Computerized Adaptive Testing Conference, Wayzata, Minnesota, June 1979.

Tukey, J.W. One degree of freedom for nonadditivity. Biometrics, 1949, 5, 232-242.

Tukey, J.W. On the comparative anatomy of transformations. Annals of Mathematical Statistics, 1957, 28, 602-632.

Whitely, S. Measuring aptitude processes with multi-component latent-trait models. Paper presented at informal meeting of ONR Contractors on model-based psychological measurement. Iowa City, Iowa, August 1980.

Navy

1 Dr. Alvah Bittner
Naval Biodynamics Laboratory
New Orleans, Lousiana 70189

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Chief of Naval Education and Training
Liason Office
Air Force Human Resource Laboratory
Flying Training Division
WILLIAMS AFB, AZ 85224

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

Navy

1 Dr. George Moeller
Head, Human Factors Dept.
Naval Submarine Medical Research Lab
Groton, CN 06340

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. William Nordbrock
Instructional Program Development
Bldg. 90
NET-PDCD
Great Lakes Naval Training Center,
IL 60088

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

Navy

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Selection and Training Research Division
Human Performance Sciences Dept.
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Worth Scanland, Director
Research, Development, Test & Evaluation
N-5
Naval Education and Training Command
NAS, Pensacola, FL 32508

1 Dr. Sam Schiflett, SY 721
Systems Engineering Test Directorate
U.S. Naval Air Test Center
Patuxent River, MD 20670

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

Navy

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MO
Brooks AFB, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 David R. Hunter
AFHRL/MOAM
Brooks AFB, TX 78235

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Marines

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 Major Michael L. Patrow, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Chief, Psychological Reserch Branch
U. S. Coast Guard (G-P-1/2/TP42)
Washington, DC 20593

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Lorraine D. Eyde
Personnel R&D Center
Office of Personnel Management of USA
1900 EStreet NW
Washington, D.C. 20415

1 Mr. Richard McKillip
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 William J. McLaurin
66610 Howie Court
Camp Springs, MD 20031

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Ina Bilodeau
Department of Psychology]
Tulane University
New Orleans, LA 70118

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. R. Darrel Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Liaison Scientists
Office of Naval Research,
Branch Office , London
Box 39 FPO New York 09510

Non Govt

1 DR. JOHN F. BROCK
Honeywell Systems & Research Center
(MN 17-2318)
2600 Ridgeway Parkway
Minneapolis, MN 55413

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr.,Fritz Drasgow
Yale School of Organization and Manageme
Yale University
Box 1A
New Haven, CT 06520

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, Texas 78228

Non Govt

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

Non Govt

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

1 Mr. Jeff Kelety
Department of Instructional Technology
University of Southern California
Los Angeles, CA 92007

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Non Govt

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Cirle, NW
Washington, DC 20036

Non Govt

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

Non Govt

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Non Govt

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. David Vale
Assessment Systems Corporation
2395 University Avenue
Suite 306
St. Paul, MN 55114

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Christopher Wickens
Department of Psychology
University of Illinois
Champaign, IL 61820

Non Govt

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2